Grayeve kode za binarne besede brez fiksnih podnizov

"Mladi za napredek Maribora" 30. srečanje

Raziskovalno področje matematike

Raziskovalna naloga

13.2.2013

Avtor: ROK HAVLAS

Mentor: SANDI KLAVŽAR

Šola: II. GIMNAZIJA MARIBOR

Contents

POVZETEK

(Binarne) besede so besede nad abecedo $0, 1$. Recimo, besede dolžine 2 so 00, 10, 01 in 11. Grayeva koda je tako zaporedje besed dolžine n, da se zaporedni besedi razlikujeta v enem mestu in da isto velja tudi za prvo in zadnjo besedo zaporedja. Na primer, zaporedje 00, 01, 11, 10 je Grayeva koda za besede dolžine 2. Grayevo kodo lahko ekvivalentno predstavimo kot hamiltonov cikel v grafu Q_n (hiperkocki). Karakterizacija podgrafov hiperkock določenih s prepovedanimi besedami, ki premorejo Grayevo kodo, je težak problem. Problem se zdi dostopnejši, če so prepovedane besede prabesede, zato sem se v raziskovalni nalogi ukvarjal z Grayevimi kodami nad besedami, kjer določen prabesedni podniz ni dovoljen. Pri tem je prabeseda taka beseda, v kateri noben njen začetek ni enak njenemu koncu iste dolžine. Ugotovil sem, da je število prepovedanih besed dolžine n rekurzivno podano

z $a_n = 2a_{n-1} + 2^{n-3} - a_{n-3}$. Prav tako sem ugotovil, da Grayeva koda ne obstaja za besede dolžine $3k$, $k \in \mathbb{N}$.

Zahvalil bi se mentorju, ki me je vodil in vzpodbujal pri zanimivem raziskovanju.

Ključne besede: binarna beseda, Grayeva koda, prabeseda, rekurzija

1 Uvod

Binarne besede so vse besede nad abecedo $0, 1$, torej, zaporedja ničel in enic. Recimo binarne besede dolžine 2 so 00, 01, 10, 11.

Naj bo n poljubno naravno število. Tedaj je Grayeva koda dolžine n tako zaporedje vseh binarnih besed dolžine n, da se zaporedni besedi razlikujeta v enem mestu (enem bitu) in da isto velja tudi za prvo in zadnjo besedo zaporedja (torej, da se prva in zadnja beseda razlikujeta v enem mestu). Grayeva koda za $n = 2$ je 00, 01, 11, 10. Grayevo kodo lahko ekvivalentno predstavimo kot hamiltonov cikel v grafu Q_n (hiperkocki).

Grayevo kodo dolžine n dobimo tako, da vzamemo Grayevo kodo dolžine $n-1$ in jo prezrcalimo (napišemo zaporedje besed v obratnem vrstnem redu) za obstoječe zaporedje in na začetku besede prvi polovici besed dodamo bit 0, drugi polovici besed pa bit 1. Za ponazoritev poglejmo, kako iz Grayeve kode za $n = 2$ dobimo Grayevo kodo za $n = 3$. Grayeva koda za $n = 2$ je

$$\begin{matrix} 0 & 0 \\ 0 & 1 \\ 1 & 1 \\ 1 & 0 \end{matrix}.$$

Ko jo prezrcalimo dobimo

$$\begin{matrix} 0 & 0 \\ 0 & 1 \\ 1 & 1 \\ 1 & 0 \end{matrix}$$

$$\begin{matrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \\ 0 & 0 \end{matrix}$$

in ko dodamo začetnice

$$\begin{matrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 0 \end{matrix}$$

$$\begin{matrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \end{matrix}$$

dobimo Grayevo kodo za $n = 3$.

Grayjevi kodi se je naprej reklo prezrcaljena binarna koda zato, ker ni imela imena in se naredi z prezrcaljenjem navadne binarne kode. Franku Grayu jo je prvi uporabil v svojem patentu leta 1953, zato je kasneje po njem tudi dobila ima Grayeva koda.

Problem ki se pojavi je, da iščemo podobne kode v podmnožicah vseh binarnih besed. Naravne podmnožice so take, kjer prepovemo fiskne podnize. Recimo če prepovemo podniz 11, potem so to Fibonaccijevi nizi. Te nize je prvi raziskoval Hsu (1993) [1], ki je uvedel tudi tako imenovane Fibonaccijeve kocke. Fibonaccijeve kocke so kocke, ki imajo v vozliščih binarne besede (kot hiperkocke), vendar nimajo ogljišč, ki bi morala vsebovala podniz 11. Te kocke so bile zelo dobro raziskane, saj je o njih znanega ogromno. Znano je, da premorejo Hamiltonove poti. Ni pa nujno, da premorejo Hamiltonove cikle, saj v primeru, če imajo liho število vozlišč cikel ni mogoč, če pa je število vozlišč sodo pa vedno premorejo cikel.

Ker je primer, ko je prepovedan podniz 11 dobro raziskan, bi pogledali bi še kakšen drug podniz. Prvi zanimiv primer je podniz 100, ki smo si ga natačno ogledali v prvem in hkrati najobsežnejšem razdelku.

Prabeseda taka beseda, v kateri noben njen začetek ni enak njenemu koncu iste dolžine. 100 je najmanjša prabeseda, ki je zanimiva - je prva

netrivialna prabeseda. Zato v četrtem razdelku pogledamo posplošitev na poljubne prabesede.

2 Grayeve kode nad besedami brez podniza 100

V tem razdelku bomo raziskovalni obstoj Grayeve kode, če je prepovedan podniz 100. V pomoč nam bodo tudi hiperkocke.

Prepovedane besede so besede, ki ne vsebujejo nekega podniza. Izberemo in prepovejmo podniz 100. Ta se prvič pojavi za $n = 3$, torej je prva prepovedana beseda 100.

$$
\begin{array}{ccc}
0 & 0 & 0 \\
0 & 0 & 1 \\
0 & 1 & 1 \\
0 & 1 & 0 \\
1 & 1 & 0 \\
1 & 1 & 1 \\
1 & 0 & 1 \\
\cancel{1} & \cancel{0} & \cancel{0}
\end{array}
$$

Za $n > 3$ dobimo ostale prepovedane besede tako, da bodisi prezrcalimo besede, ki so bile prepovedane že prej ali tako, da besedam dodamo začetnico 1, ko delamo besede dolžine $n + 1$.

Poglejmo, koliko besed je prepovedanih v Grayevi kodi dolžine n. Naj bo a_n število prepovedanih besed (besed, ki vsebujejo podniz 100) dolžine n.

Najprej določimo, koliko besed ima podniz 100 na začetku besede. Od vseh 2^n besed dolžine n jih ima polovica - 2^{n-1} besed prvi bit 1. Od vseh teh besed, ki imajo prvo bit 1, jih ima polovica drugi bit 0 (ker smo drugi bit dobili tako, da smo prezrcalili binarne besede dolžine $n - 1$ - polovica besed

je tam imela 0 za prvi bit), tako da se skupno 2^{n-2} besed začne z 10. Med besedami ki se začnejo z 10, jih je polovica takih, ki imajo bit 0 na 3. mestu, tako da je vse skupaj 2^{n-3} binarnih besed, ki se začnejo s 100.

Binarne besede ki nimajo podniza 100 na začetku besede, temveč na sredini ali na koncu besede so bile dobljene z zrcaljenjem Grayeve koda za $n-1$. Ker je a_{n-1} prepovedanih besed v Grayevi kodi za $n-1$ in mi to kodo prezrcalimo, dobimo $2a_{n-1}$ prepovedanih besed, v katerih podniz 100 nastopa v sredini ali na koncu besede.

Ampak v tem primeru smo preveč krat šteli besede, ki imajo podniz 100 tako na začetku kot tudi v sredini/na koncu besede. Ker so takšne besede bile prepovedane tudi brez podniza 100 na začetku besede, ko je bila njihova dolžina $n-3$, so bile vse prepovedane besede dolžine $n-3$ štete dvakrat, zato jih moremo odšteti. Takšnih besed je a_{n-3}.

Tako lahko zapišemo število prepovedanih besed dolžine n z rekurzivno zvezo:

$$a_n = 2a_{n-1} + 2^{n-3} - a_{n-3}\,.$$

Homogen del enačbe je

$$z_n - 2z_{n-1} + z_{n-3} = 0\,.$$

Temu lahko priredimo karakteristični polinom

$$Q(x) = x^3 - 2x^2 + 1 = 0$$

in najdemo njegove ničle. Ničle karakterističnega polinoma so

$$x_1 = 1, x_2 = \frac{1+\sqrt{5}}{2}, x_3 = \frac{1-\sqrt{5}}{2}\,.$$

Za $n = 1$ ni prepovedanih besed, zato $a_1 = 0$. Tudi če je $n = 2$ ni prepovedanih besed, zato $a_2 = 0$. Za $n = 3$ je prepovedana beseda 100, zato

$a_3 = 1$. To so torej naši začetni pogoji.

Za homogeni del velja

$$z_n = A \cdot 1^n + B \cdot (\frac{1+\sqrt{5}}{2})^n + C \cdot (\frac{1-\sqrt{5}}{2})^n.$$

Nehomogeni del enačbe je

$$b_n = D \cdot 2^n.$$

Če to vstavimo v rekurzivno enačbo, dobimo

$$D \cdot 2^n = 2D \cdot 2^{n-1} + 2^{n-3} - D \cdot 2^{n-3}$$

od koder z deljenjem z 2^{n-3} dobimo enačbo

$$8D = 8D + 1 - D.$$

Zaključimo lahko, da je $D = 1$. Splošna rešitev rekurzivne enačbe je vsota posebne homogene rešitve in nehomogene rešitve, torej

$$a_n = z_n + b_n,$$

$$a_n = A \cdot 1^n + B \cdot (\frac{1+\sqrt{5}}{2})^n + C \cdot (\frac{1-\sqrt{5}}{2})^n + 2^n\,.$$

če vstavimo tri začetne pogoje dobimo tri enačbe s tremi neznankami:

$$0 = A \cdot 1^1 + B \cdot (\frac{1+\sqrt{5}}{2})^1 + C \cdot (\frac{1-\sqrt{5}}{2})^1 + 2$$

$$0 = A \cdot 1^2 + B \cdot (\frac{1+\sqrt{5}}{2})^2 + C \cdot (\frac{1-\sqrt{5}}{2})^2 + 4$$

$$1 = A \cdot 1^3 + B \cdot (\frac{1+\sqrt{5}}{2})^3 + C \cdot (\frac{1-\sqrt{5}}{2})^3 + 8$$

od koder dobimo $A = 1$, $B = -1 - \frac{2\cdot\sqrt{5}}{5}$, $C = -1 + \frac{2\cdot\sqrt{5}}{5}$ in splošna rešitev se tako glasi

$$a_n = 1 - (1 + \frac{2\cdot\sqrt{5}}{5}) \cdot (\frac{1+\sqrt{5}}{2})^n + (-1 + \frac{2\cdot\sqrt{5}}{5}) \cdot (\frac{1-\sqrt{5}}{2})^n + 2^n\,.$$

2.1 Pogled skozi hiperkocke

Hamiltonov cikel na hiperkocki obstaja, če obstaja takšna pot, ki gre skozi vsako vozlišče natanko enkrat in se začne in zaključi v isti točki. Hamiltonva pot na hiperkocki je takšna pot, ki začne v enem vozlišče in gre skozi vsa ostala vozlišča. Obstoj Hamiltonovega cikla na hiperkocki je ekvivalenten obstoji Grayeve kode, saj gre Hamiltonov cikel skozi vsako vozlišče, ki pa je na hiperkocki ravno binarna beseda. Hiperkocko Q_n lahko tudi obravnavamo kot dvodelni graf. V dvodelnem grafu Hamiltonov cikel obstaja če in samo če je število elementov v disjunktnih množicah enako, torej če je skupno število elementov sodo. Torej za sodi a_n, Hamiltonov cikel lahko obstaja, za lihi a_n pa Hamiltonov cikel ne obstaja. Ker je a_n dan z enačbo

$$a_n = 2a_{n-1} + 2^{n-3} - a_{n-3}$$

in ker je $2a_{n-1}$ vedno sodo, ker je 2 sodo in je tudi 2^{n-3} sodo za vse $n > 3$, moramo pogledati, ali je a_{n-3} sod ali lih. če bo a_{n-3} sod/lih, bo tudi a_n enake parnosti, torej je parnost členov za indekse po modulu 3 invarnianta. Torej, če bo nek a_n sod, bo tudi a_{n+3} sod in posledično tudi bodo tudi a_{n+6},...a_{n+3k}; $k \in \mathbb{N}$ sodi. če pa bo a_n lih, bo tudi a_{n+3} lih in posledično tudi bodo tudi a_{n+6},...a_{n+3k} lihi.

Zdaj si bomo pogledali kaj je s Hamiltonovim ciklom.

Poglejmo kaj se dogaja z a_4. Z uporabo rekurzivne enačbe dobimo

$$a_4 = 2a_3 + 2^1 - a_1 = 2 + 2 - 0 = 4\,.$$

Ker je a_1 sod je tudi a_4 sod in vsak a_n, $n = 3k + 1$; $k \in \mathbb{N}$, bo sod, tako da Hamiltonov cikel lahko obstaja za take n.

Izkaže se, da za dovoljene besede dolžine 4 res obstaja Hamiltonov cikel:

$$\begin{matrix} 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{matrix} .$$

Če za a_5 uporabimo rekurzivno enačbo dobimo

$$a_5 = 2a_4 + 2^2 - a_2 = 8 + 4 - 0 = 12 \,.$$

Ker je a_2 sod je tudi a_5 sod in vsak a_n, $n = 3k + 2$; $k \in \mathbb{N}$ bo sod, tako da bi Hamiltonov cikel za tak n lahko obstajal.
Tudi za dovoljene besede dolžine $n = 5$ obstaja Hamiltonov cikel:

$$
\begin{array}{ccccc}
0 & 0 & 1 & 0 & 1 \\
0 & 1 & 1 & 0 & 1 \\
0 & 1 & 1 & 1 & 1 \\
0 & 1 & 1 & 1 & 0 \\
0 & 1 & 0 & 1 & 0 \\
0 & 1 & 0 & 1 & 1 \\
1 & 1 & 0 & 1 & 1 \\
1 & 1 & 0 & 1 & 0 \\
1 & 1 & 1 & 1 & 0 \\
1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 0 & 1 \\
1 & 0 & 1 & 0 & 1 \\
1 & 0 & 1 & 1 & 1 \\
1 & 0 & 1 & 1 & 0 \\
0 & 0 & 1 & 1 & 0 \\
0 & 0 & 1 & 1 & 1 \\
0 & 0 & 0 & 1 & 1 \\
0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1
\end{array}
.
$$

Sedaj pa bomo dokazali, da Hamiltonov cikel ne more obstajati za $n = 3k$; $k \in \mathbb{N}$.

Če za a_6 uporabimo rekurzivno enačbo, dobimo

$$a_6 = 2a_5 + 2^3 - a_3 = 24 + 8 - 1 = 31\,.$$

Ker je a_3 lih je tudi a_6 lih in vsak a_n, $n = 3k$ bo lih, tako da Hamiltonov cikel za tak n ne obstaja.

Tako smo ugotovili, da Hamiltonov cikel lahko obstaja za $n = 3k + 1$ ali $n = 3k+2$ (za $n = 4$ in $n = 5$ obstaja), za $n = 3k$ pa ne obstaja in posledično tudi Grayeva koda ne obstaja. Na podlagi obstoja Hamiltonovega cikla za $n = 4$ in $n = 5$ domnevam, da Hamiltonov cikel obstaja za vse $n = 3k + 1$ in $n = 3k + 2$.

Pa poglejmo še, kako je s Hamiltonovo potjo.

Za $n = 3$ vidimo, da ob izločitvi vseh prepovedanih besede še vedno obstaja Hamiltonovo pot od prvega člena do zadnjega dovoljenega člena. Kaj pa se zgodi, če pogledamo za $n = 4$? Grayevo kodo za $n = 4$ dobimo z zrcaljenjem Grayeve kode za $n = 3$ in dodajanjem predpon 0 ali 1. Ker se zadnja beseda Grayeve kode za n preslike takoj za samo sabo, obstaja med njima Hamiltonova pot, saj se razlikujeta samo v predponi. če pa je zadnja beseda Grayeve kode za n prepovedana, je tudi prva prezrcaljena beseda prepovedana, tako da se obe "odstranita". Potem sta predzadnja beseda Grayeve kode za n in beseda, ki je njena zrcalna slika razlikujeta samo v predponi in obstaja med njima Hamiltonova pot. Tako je tudi v našem primeru, saj je bila prva prepovedana beseda 100, ki pa je na zadnjem mestu zaporedja besed pri Grayevi kodi za $n = 3$. Ko zaporedje prezrcalnimo torej obstaja Hamiltonova pot za novo nastalo zaporedje. Vsi primeri kjer je podniz 100 v sredini oz. na koncu besede nastanejo z zrcaljenjem predhodnega zaporedja. Ker velja za predhodno zaporedje, da prepovedana člena na sredini ali koncu besede ne motita Hamiltonove poti, to velja tudi za novo zaporedje, saj tako oba para členov (člen pred prepovedano besedo in člen za prepovedano besedo) dobita iste predpone. Besede, ki pa dobijo komaj v zadnjem koraku predpono 100 in še prej niso bile prepovedane pa se nahajajo na zadnjih mestih zaporedja, saj so nastale kot zrcalne slike besed z začetnico 00 iz prejšnjega zaporedja, ki pa se nahajajo vedno na začetku zaporedja (saj vedno tistim prvim členom dodajamo predpono 0. Tako da te besede ne ovirajo Hamiltonove poti. Zaradi tega vedno obstaja Hamiltonova pot od prvega člena zaporedja do zadnjega dovoljenega člena zaporedja.

3 Algoritem za iskanje Grayeve kode

Recimo, da Grayevo kodo razdelimo na dele. 2^n-ti del naj obsegajo prvih $2^{n-1} + 1$ do 2^n besed.

Ker število besed narašča eksponentno in je za velike n-je težko pisati

vse besede predlagam za iskanje Grayeve kode naslednji algoritem: Zadnjo dovoljeno besedo povežimo z njeno zrcalno sliko (besedo z različno predpono in enakim ostalim delom). Ker je prepovedanih zadnjih 2^{n-3} besed, bo dana zrcalna slika v $2^{n-3+1} = 2^{n-2}$ delu. Zadnjo besedo iz 2^{n-2} dela povežimo z vsemi nadaljnimi členi zaporedja do vključno z zadnjim dovoljenim členom. Tako je naša naloga samo še najti Hamiltonovo pot med prvimi 2^{n-2} dovoljenimi besedami. Za začetek in konec naj ima zadnjo besedo 2^{n-2}-ega dela, za konec pa zrcalni člen (člen ki se razlikuje od originalnega samo v predponi) zadnjega dovoljenega člena zaporedja.

4 Grayeve kode nad besedami, ki ne vsebujejo fiksne prabesede

V tem razdelku bomo poskusili posplošiti rezultate, ki smo jih dobili za prepovedan podniz 100 poljuben prepovedan podniz.

Težave imamo pri posplošitvi besedah, ki imajo nek začeten del enak nekemu končnemu delu, recimo 110011 ali 1101, saj prehod med koncem ene besede in začetkom druge besede ni jasen - ena beseda se lahko nadaljuje v drugo, na primer 1100110011 ali 1101101, česar pri 100 ni bilo. Zato bi se omejili na besede, katerih začetek ni enak koncu. Takšnim besedam rečemo prabesede. Torej prabeseda je taka beseda, v kateri noben njen začetek ni enak njenemu koncu iste dolžine.

Spet naj bo a_n število prepovedanih besed za Grayevo kodo dolžine n. Podniz naj bo dolžine l; $l \in \mathbb{N}$. število prepovedanih besed se po v vsakem naslednjem koraku Grayeve kode podvojilo, saj se bo podvojilo število besed, ki imajo dan podniz v sredini/na koncu besede. Takih je torej $2a_{n-1}$. Število prepovedanih besed, ki imajo prepovedan podniz na začetku besede bo 2^{n-l}, saj teh l črk mora biti na začetku besede, za vsako od ostalih $n - l$ mest pa imamo 2 možnosti, saj lahko na vsako izmed teh mest postavimo bodisi 0 bodisi 1, kar ne vpliva na to, katere črke so na začetku (ker delamo z

prabesedami). Vendar je med temi 2^{n-l} prepovedanimi besedami tudi a_{n-l} takšnih, ki imajo prepovedan podniz tudi na sredini/na koncu besede, torej, ki so bile prepovedane že pred l koraki, preden smo dodali teh l začetnih črk.

Torej je število prepovedanih besed dolžine n podano z rekurzivno zvezo

$$a_n = 2a_{n-1} + 2^{n-l} - a_{n-l}.$$

Karakterističnega polinoma tukaj ne moremo nastaviti, ker je stopnja lahko poljubno velika (saj je lahko l poljubno velik). Opazimo pa lahko nekaj druga. Prva prepovedana beseda se pojavi za $n = l$ (saj ima l črk). Torej je za $n = l$ prepovedana samo ena beseda. Ker sta spet $2a_{n-1}$ in 2^{n-l} soda za $n > l$, bo parnost a_n odvisna le od člena a_{n-l}. Za $n = 2l$ bo torej a_{2l} odvisen od a_l ki je lih, torej bo tudi a_{2l} lih in vsak $n = kl$; $k \in \mathbb{N}$ bo lih, torej za $n = kl$ Hamiltonov cikel in posledično tudi Grayeva koda ne obstaja. Za prabesede dolžine l lahko uporabljamo tudi predlagan algoritem, samo da $n - 3$ zamenjamo z $n - l$.

5 Zaključek

V nalogi sem uspel dokažati, da za nekatere n-je ne obstajajo Grayeve kode s prepovedanimi prabesednimi podnizi. Zaradi velikega števila možnih besed pa nisem uspel končno dokazati obstoja Grayeve kode za ostale n-je. Zaradi tega sem se domislil algoritma, ki problem bistveno skrči. Vendar ker ne vemo, ali vedno obstaja Hamiltonova pot med danimi besedami, algoritem ni čisto dober. Dober znak pa je, da med dovoljenimi besedami dolžine n vedno obstaja Hamiltonova pot. Dober znak je tudi, da za prepovedan podniz 100 obstaja Grayeva koda za $n = 4$ in $n = 5$. Zaradi obsežnosti in težvanosti problema sem s svojimi rezultati zadovoljen, vendar bom v bližnji prihodnosti kljub temu poskusil najti kakšno izboljšavo, predvsem algoritma.

6 Literatura

[1] W.-J. Hsu, Fibonacci cubes - a new interconnection technology, IEEE Trans. Parallel Distrib. Syst. 4 (1993) 3-12.